Provenance in RDF

Jeni Tennison

Overview

- Why is it important?
- What do we need to know?
- How can it be represented?
- Where from here?

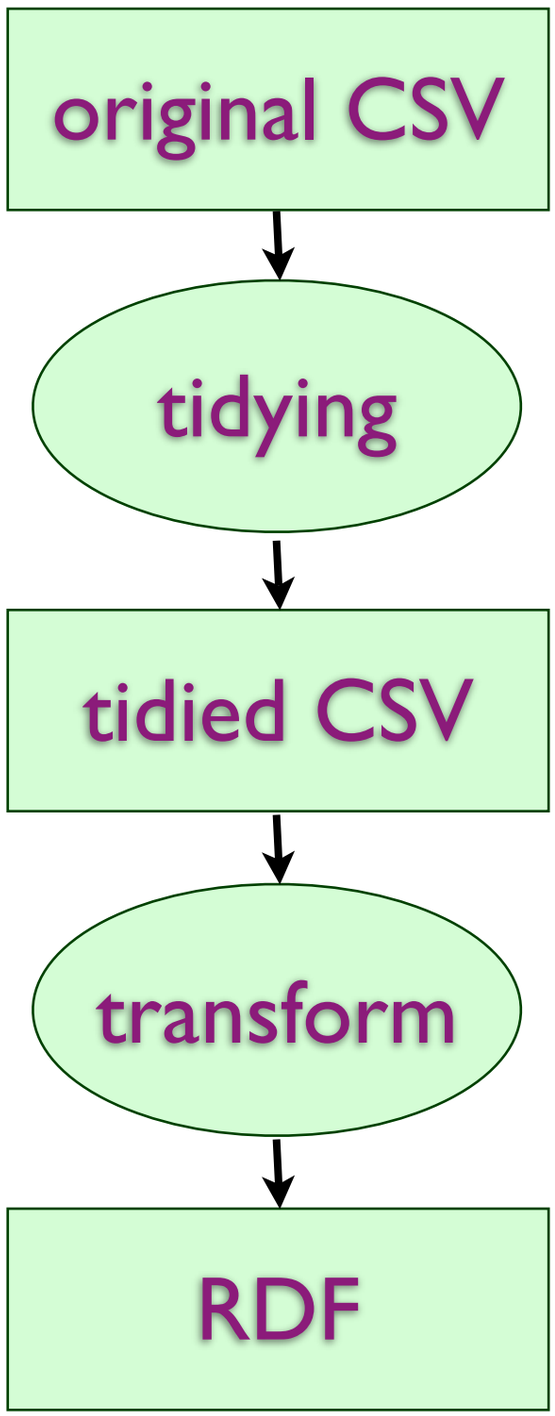
Provenance -> Trust

- Paranoid questions
  - where's the data from? when?
  - what's been done to it? when?
  - who did it? when?
- Can I replicate it?
  - Open Data + Open Source

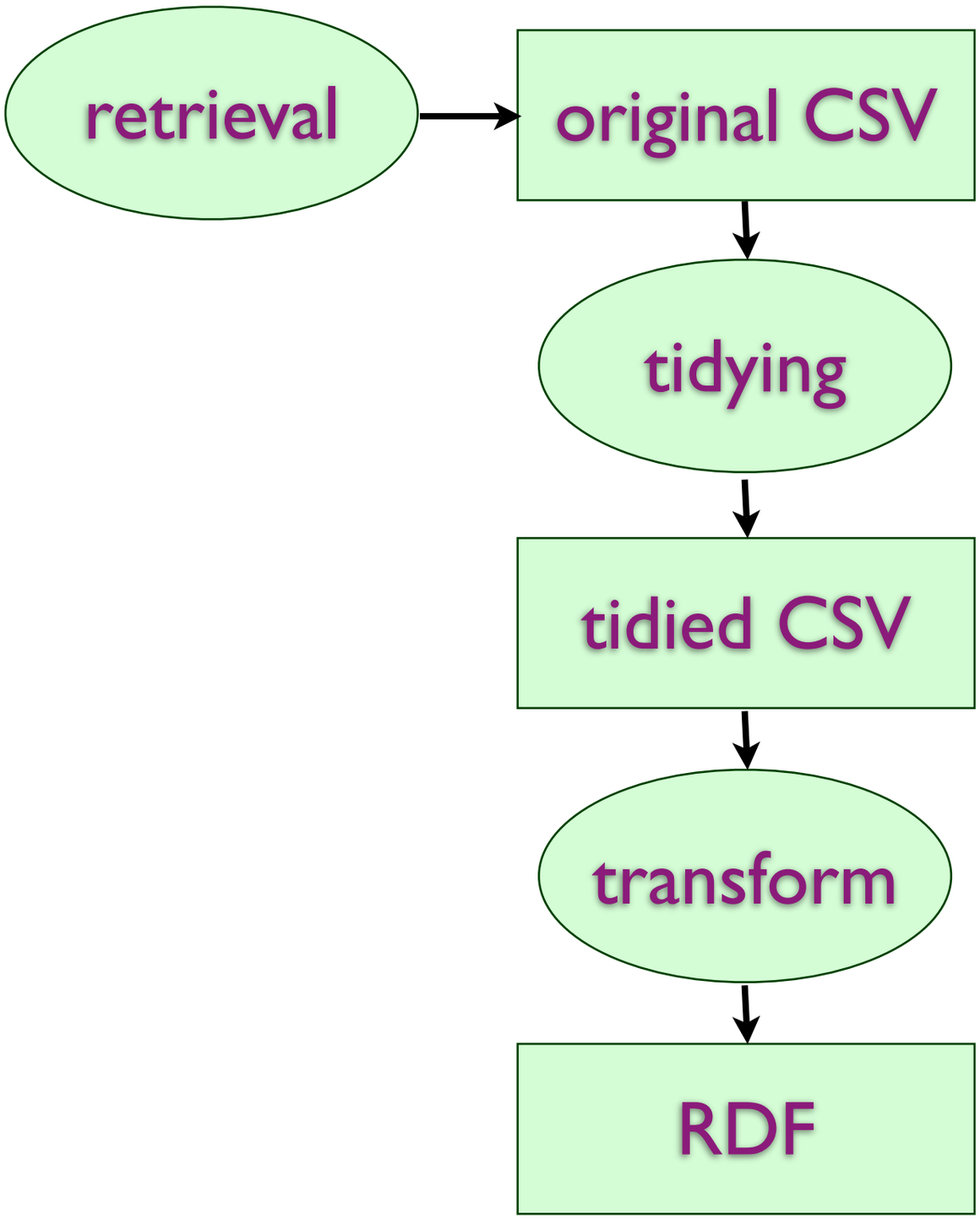
Eg Traffic Flow Data

RDF

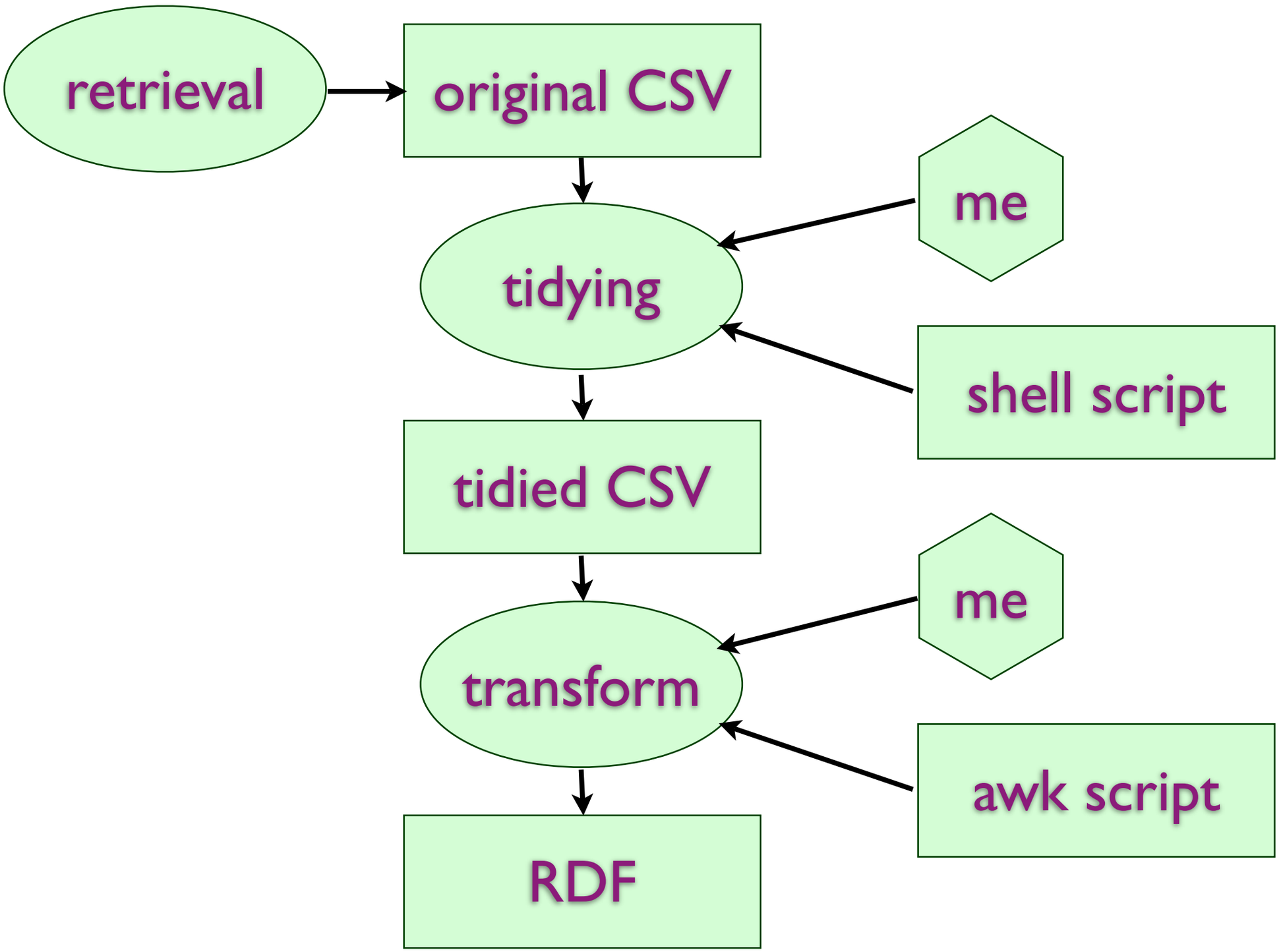
Eg Traffic Flow Data



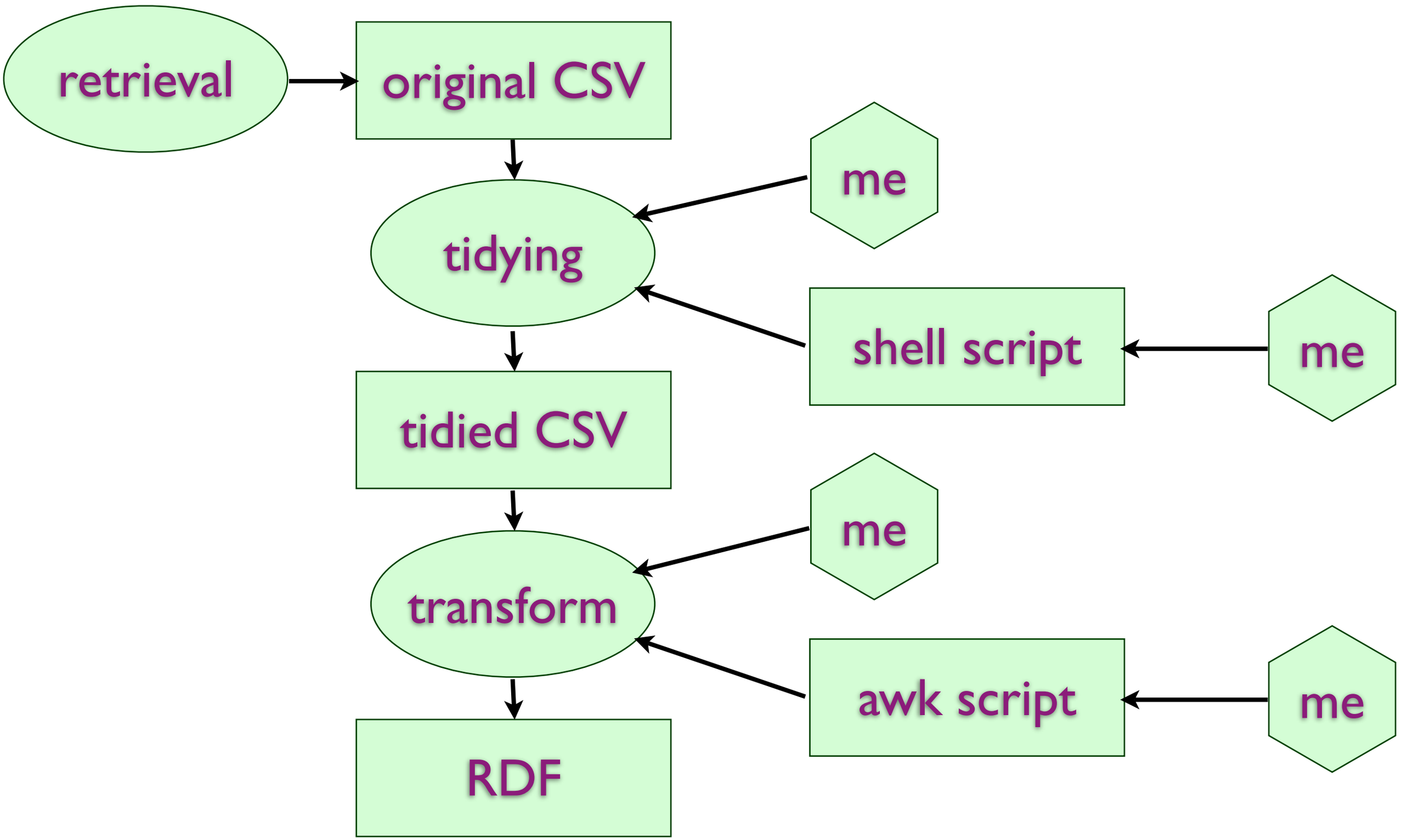
Eg Traffic Flow Data



Eg Traffic Flow Data



Eg Traffic Flow Data



Different Levels

- Legislation & sections
  - created by parliament, legislation
- Legislation content
  - amended by other legislation
- Marked up documents
  - automated, on-demand etc

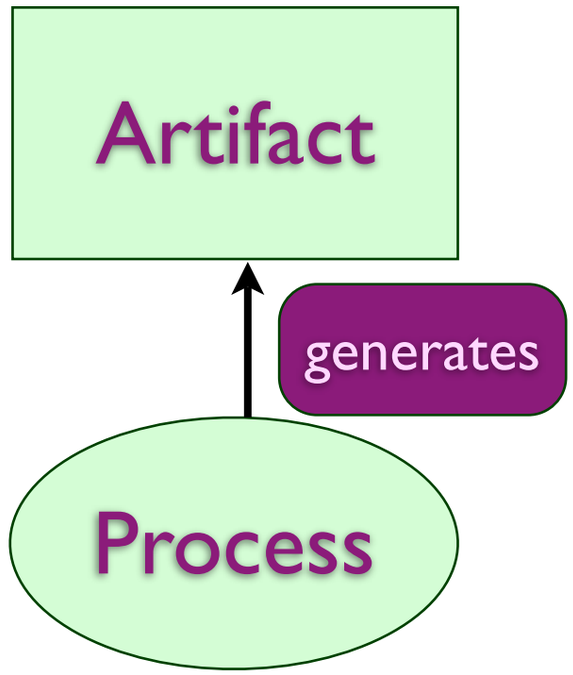
Open Provenance Model

- http://openprovenance.org/
- International effort
- Technology independent
  - XML & RDF representations
- Covers a range of use cases
  - cars, food, data

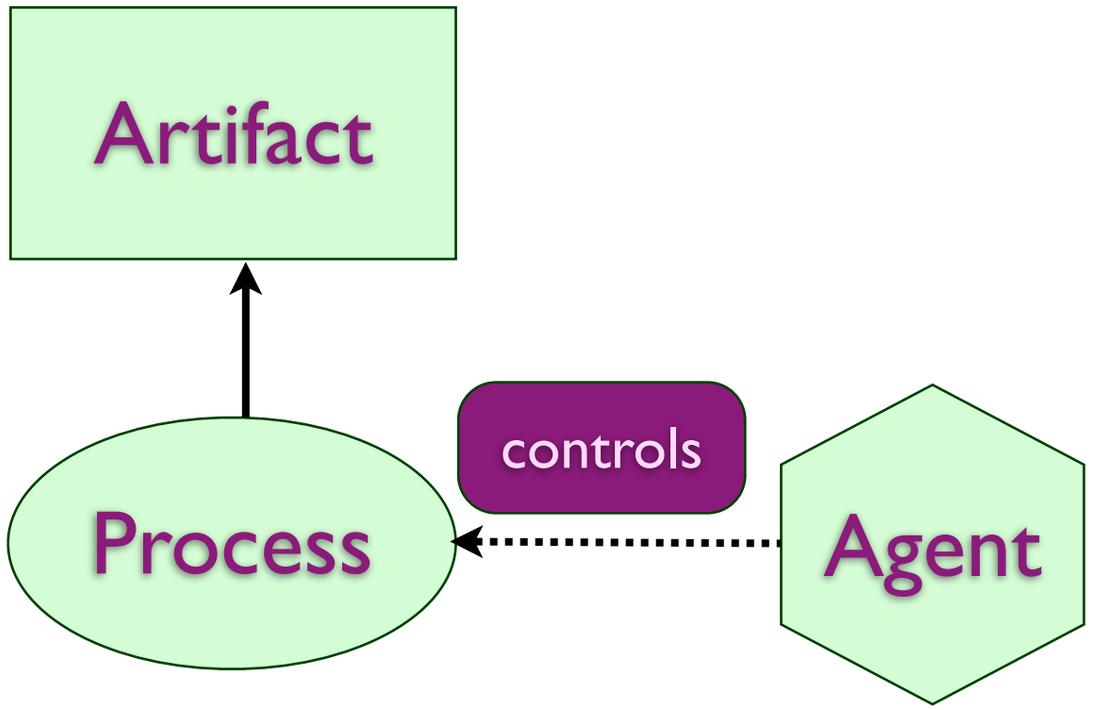
Open Provenance Model

Artifact

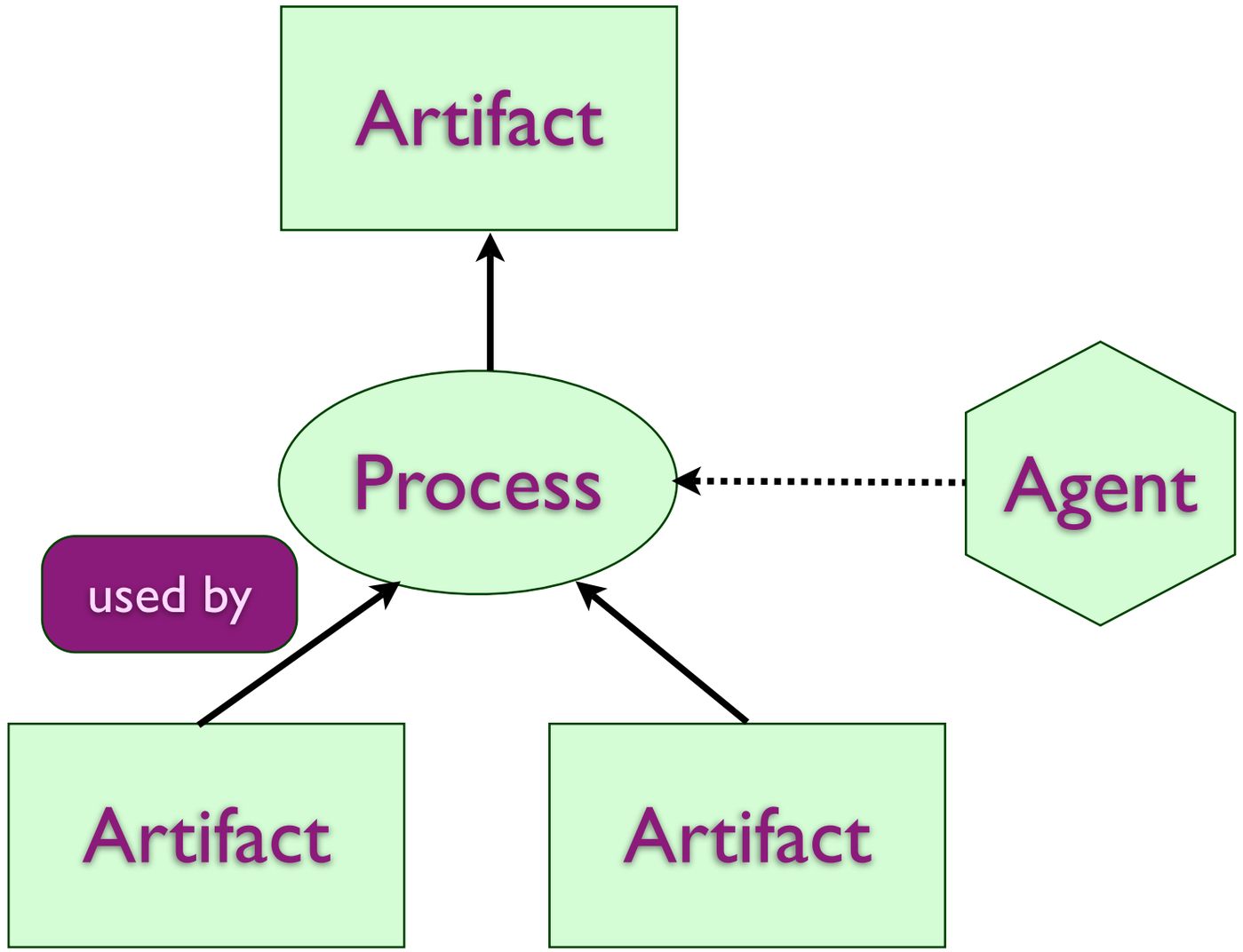
Open Provenance Model



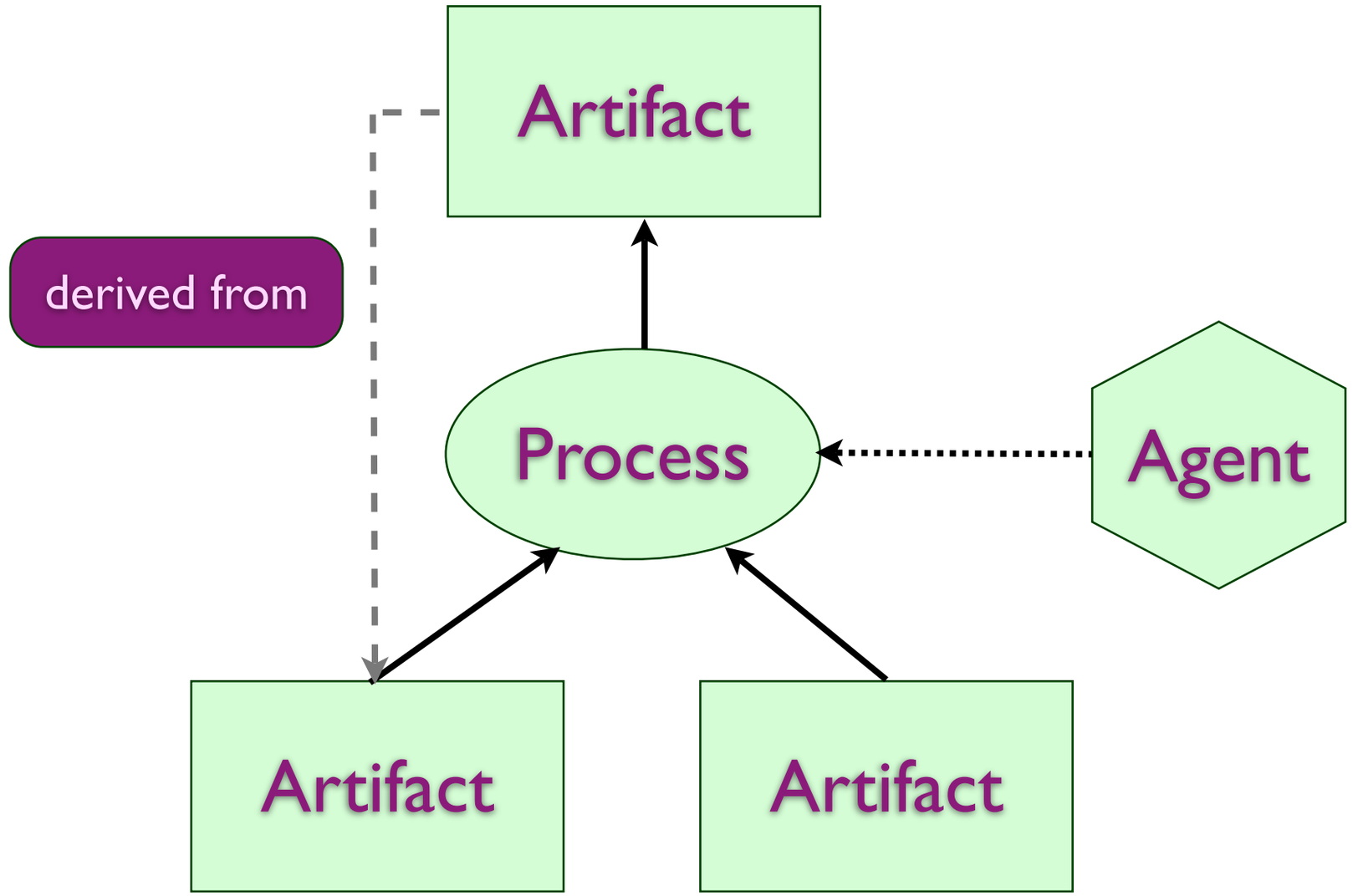
Open Provenance Model



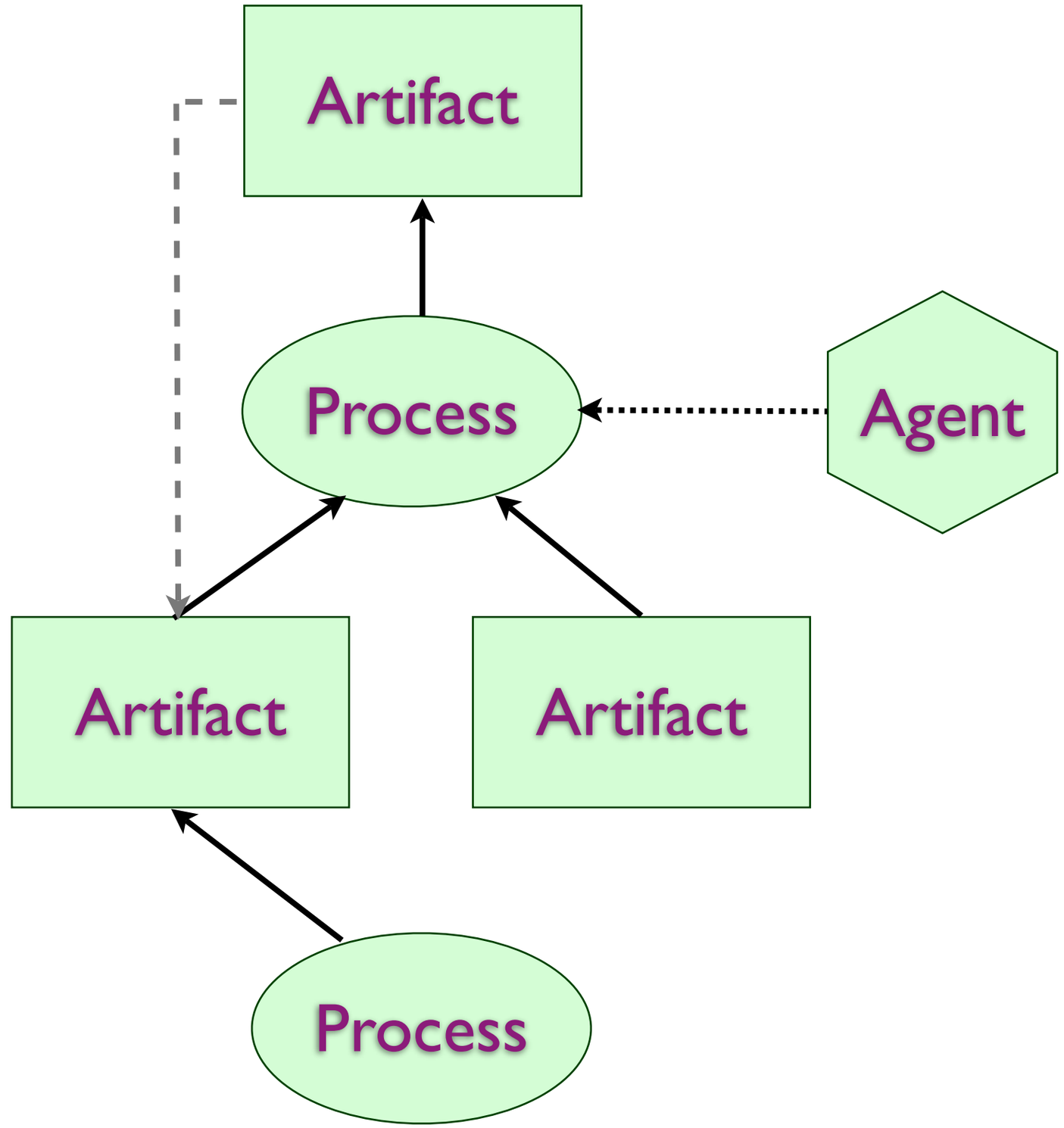
Open Provenance Model



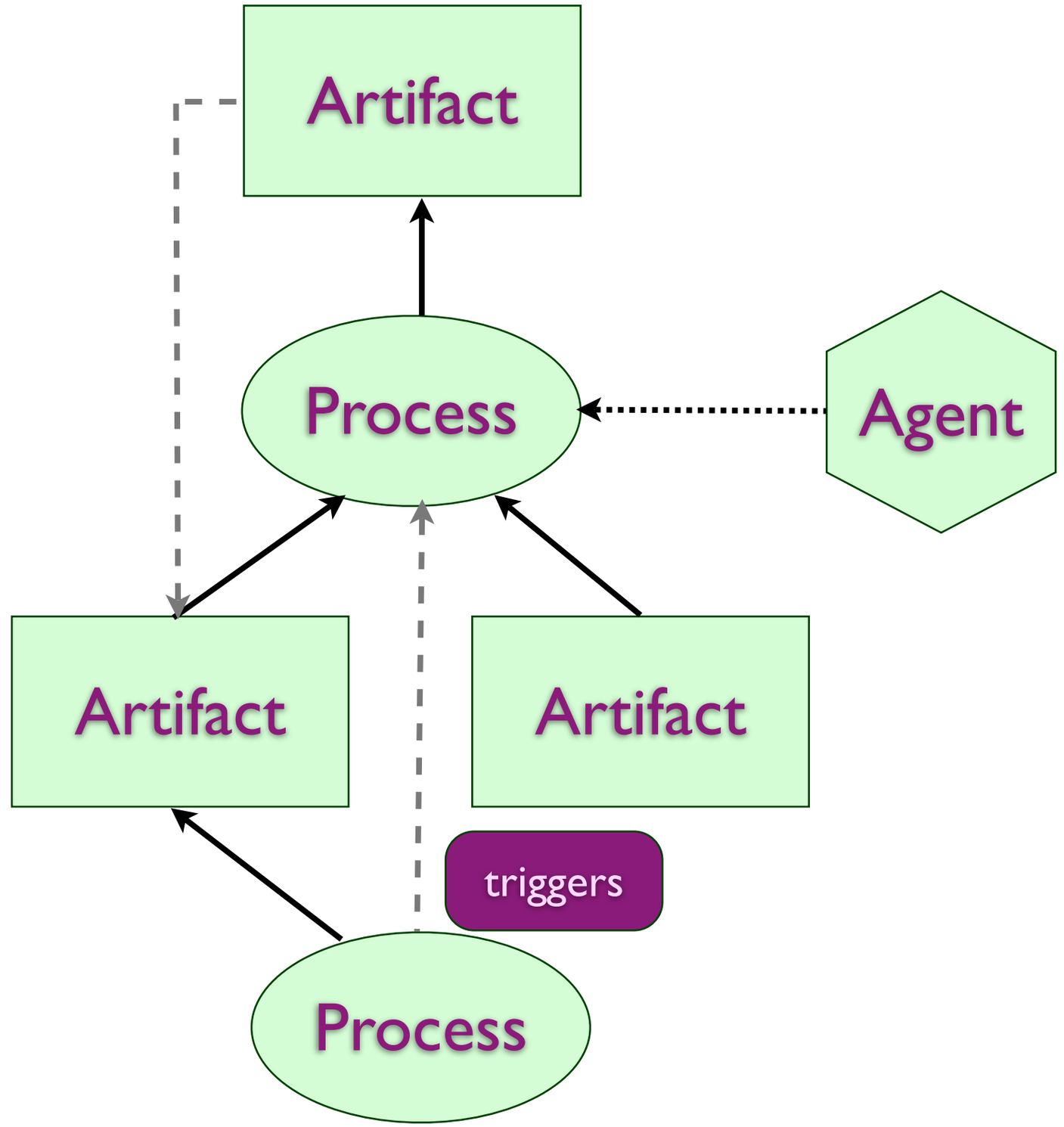
Open Provenance Model



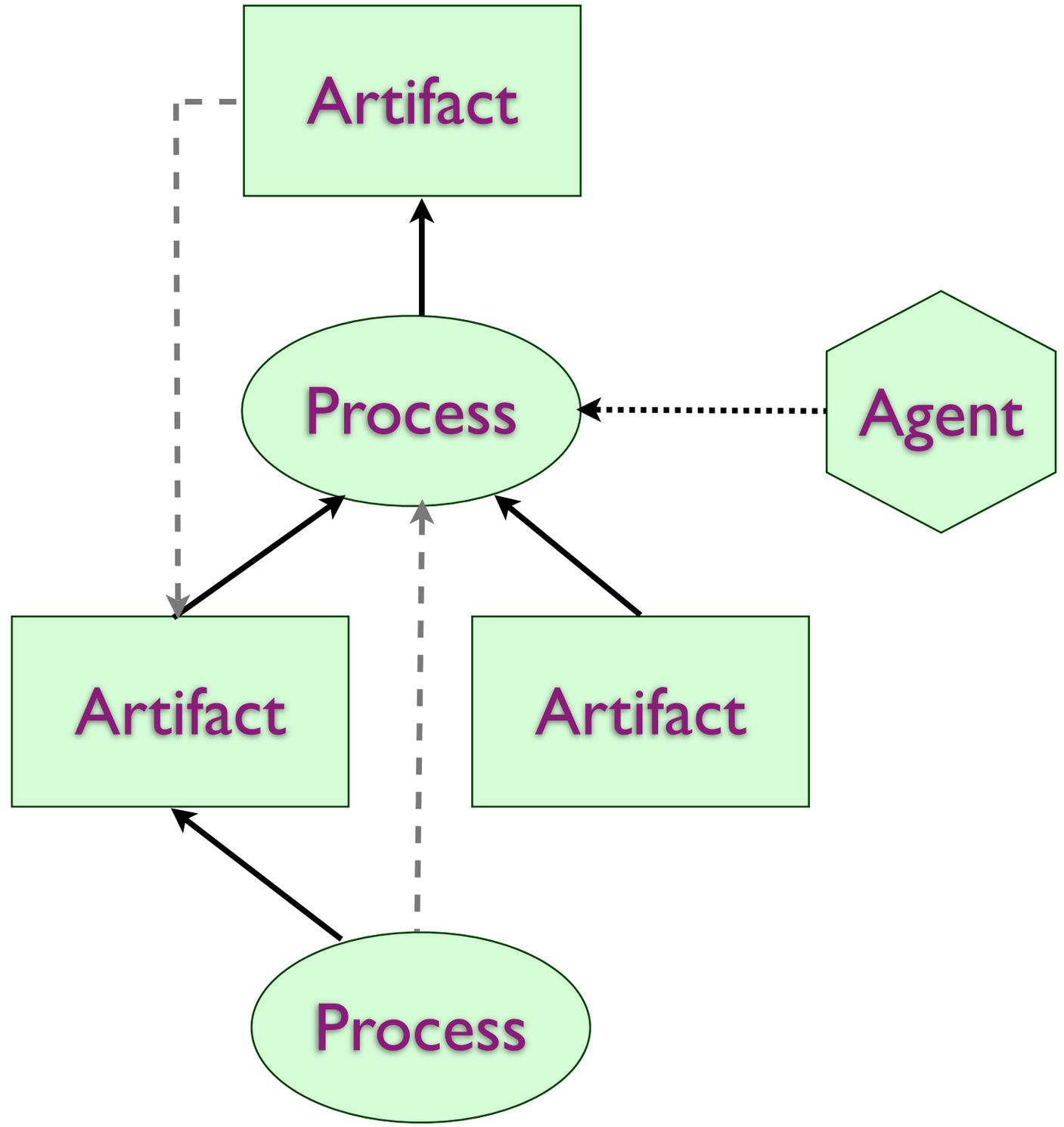
Open Provenance Model



Open Provenance Model



Open Provenance Model



Complications

- Roles for relationships
  - could be done with sub-properties
- Timestamps on relations
  - (rather than on processes)
  - requires n-ary relations
- Multiple accounts

Provenance Vocabulary

- http://purl.org/net/provenance/ns
  - Olaf Hartig & Jun Zhao
  - work in progress
- Provenance for RDF
  - focus on RDF graphs/datasets

Provenance of what?

- Files?
- Datasets? (VoID)
- Graphs?
  - require named graphs
- Triples?
  - require reification

Where From Here?

- Need a vocabulary
  - shape Provenance Vocabulary?
  - simplified OPM?
- Need to name statements
  - named graphs?
- Need for tool support

Questions?